

Revista de Instrumentos, Modelos e Políticas em Avaliação Educacional **IMPA**

Teoria clássica dos testes e teoria de resposta ao item em avaliação educacional

Leandro Araujo de Sousaⁱ 

Instituto Federal Ceará (IFCE), Canindé, Ceará, Brasil

Adriana Eufrásio Bragaⁱⁱ 

Universidade Federal do Ceará, Fortaleza, Ceará, Brasil

Resumo

Este artigo caracteriza-se como um ensaio teórico que aborda aspectos introdutórios da psicometria clássica e moderna. Dessa forma, apresentamos um breve histórico do desenvolvimento, assim como conceitos e características gerais de cada uma das abordagens. Para a teoria clássica dos testes apresentamos seu modelo básico, pressupostos e os parâmetros dos itens comumente utilizados, a dificuldade e discriminação. Para a teoria de resposta ao item apresentamos o conceito de curva característica do item, os parâmetros frequentemente utilizados, a dificuldade, discriminação e acerto casual, os pressupostos e os modelos mais comuns. Com este ensaio, esperamos apresentar os conceitos mais básicos para os iniciantes no estudo da psicometria no âmbito da avaliação educacional.

Palavras-chave

Psicometria. Teoria clássica das medidas. Teoria de resposta ao item.

Classical test theory and item response theory in educational assessment

Abstract

This article is characterized as a theoretical essay that addresses introductory aspects of classical and modern psychometry. Thus, we present a brief history of development, as well as concepts and general characteristics of each of the approaches. For the classic test theory we present its basic model, assumptions and the parameters of the commonly used items, the difficulty and discrimination. For the item response theory, we present the concept of the item's characteristic curve, the frequently used parameters, the difficulty, discrimination and casual adjustment, the most common assumptions and models. With this essay, we hope to present the most basic concepts for beginners in the study of psychometry within the scope of educational assessment.

Keywords

Psycometry. Classical test theory. Item response theory.

1 Introdução

A psicometria iniciou o seu desenvolvimento no século XIX, junto com a consolidação da ciência com base no positivismo, com a necessidade de elaboração de instrumentos de avaliação psicológica que proporcionassem medidas válidas e precisas (SARTES; SOUSA-FORMIGONI, 2013). Tem fundamento em uma base epistemológica eminentemente quantitativista, assumindo pressupostos da teoria da

mensuração, e é considerada como um ramo da psicologia, ou seja, das ciências empíricas (PASQUALI, 2009). Tem o objetivo de realizar medidas objetivas dos aspectos mentais. Pode ser definida como um conjunto de métodos e técnicas que utilizam de parâmetros métricos indispensáveis à medição das variáveis psicológicas independentes do campo de aplicação e dos instrumentos empregados (MUÑIZ, 1994).

Arias, Lloreda, Lloreda (2014) afirmam ser consenso, embora existam críticas, que a psicometria como teoria da medição pode ser considerada como o processo de atribuição de números aos atributos dos sujeitos de modo a revelar seus diferentes graus. É importante ressaltar que a psicometria é um termo bastante amplo e se refere a pelo menos cinco áreas especializadas, a saber: teoria da medição, teoria dos testes, escalas psicológicas, escalas psicofísicas e técnicas multivariadas (MUÑIZ, 1994). Neste texto será tratada, por especificidade, da Teoria dos Testes.

Inicialmente a teoria dos testes fundamentou-se unicamente no enfoque clássico, a teoria clássica dos testes (TCT). No entanto, com os estudos que buscaram definir a estrutura dos traços latentes como causadores do comportamento observável surge a TRI (SARTES; SOUSA-FORMIGONI, 2013).

A TCT predominou por muito tempo, até meados da década de 1980. Muito por conta do seu modelo matemática relativamente simples. A teoria de resposta ao item (TRI), por outro lado, possui um modelo matemático mais sofisticado e, embora tenha surgido bem antes, só começou a ser efetivamente utilizada na validação de instrumentos e análise dos resultados a partir dos anos de 1980 com o avanço tecnológico que permitiu a criação de programas de computadores que possibilitassem sua implementação (SARTES; SOUSA-FORMIGONI, 2013).

Desde então, a TRI tem sido amplamente utilizada em testes educacionais padronizados e em larga escala, como o Enem (SOUSA; BRAGA; TROMPIERI FILHO, 2015), assim como em outras áreas (MOREIRA JUNIOR, 2010; SOUSA et al., 2018). Diante disso, este ensaio objetiva apresentar brevemente conceitos e características da teoria clássica dos testes e teoria de resposta ao item mais característicos do campo da avaliação educacional. Dessa forma, enfatizamos conceitos da TCT e da TRI, como os parâmetros dos itens e modelos mais comuns

que são aplicados na área de avaliação educacional, mais precisamente a testes de desempenho escolar.

2 Breve história e desenvolvimento da teoria dos testes

O desenvolvimento dos testes, ao considerar uma perspectiva histórica, tem seu início em tempos bem remotos, desde a Antiguidade. Registra-se que já em 3.000 a.C o Império Chinês utilizava-se de testes para selecionar bons soldados para compor o exército (URBINA, 2004).

A teoria dos testes surgiu no início do século XX, inicialmente com os trabalhos de Spearman em 1904, ao propor o modelo linear clássico, na tentativa de fundamentação dos escores e da estimação dos erros de medidas associados em um teste (MUÑIZ, 1994; TRAUB, 1997), também teve contribuições de George Udny Yule, Truman Lee Kelley e outros (TRAUB, 1997). Esse modelo tornou-se o pressuposto fundamental da TCT, visto adiante com mais detalhes.

Alfred Binet e Théodore Simon tiveram papel de “pai fundador” da TRI (LINDEN, 2015). Binet foi solicitado a desenvolver na cidade de Paris um teste capaz de diferenciar estudantes com retardo mental e direcioná-los para a educação especial. A partir, disso, Binet pensou formas de medir uma variável não observável, do qual não se tem acesso direto, o que hoje denominamos de traço latente, variável latente ou teta. Em 1905, apenas um ano após a publicação de Spearman em 1904, Binet publica um trabalho em que estão explícitas essas ideias, ou seja, os modelos de TCT e TRI surgiram no mesmo período.

Gulliksen, na década de quarenta apresentou uma sistematização da TCT, mostrando toda a sua estrutura, em que apresentou os postulados do modelo clássico detalhadamente (GULLIKSEN, 1943).

Também na mesma década Stevens (1946), ao apresentar as escalas de medidas, indica uma solução para os problemas referentes à mensuração de sensações humanas até então elencadas pela *Committee of the British Association for the Advancement of Scienc.* Entre os principais problemas era a própria definição de medição. Com isso, Stevens propôs a definição de medida sob uma variedade de formas, que estão associadas às propriedades da operação empírica com o objeto e

às propriedades matemáticas das escalas. Dessa forma, as análises estatísticas empregadas deveriam levar em consideração a natureza da escala em que o objeto está sendo medido (STEVENS, 1946).

Thurstone também apresentou importantes contribuições para a teoria dos testes, principalmente ao publicar seu livro sobre análise fatorial, proporcionando grandes avanços na verificação da validade dos testes, sendo potencializados com a implementação de recursos computacionais (MUÑIZ, 1994). Thurstone também contribuiu para a TRI. Diferentemente de Binet, ele despreendeu a inteligência da idade e construiu uma escala própria, impondo uma curva conhecida, a curva de distribuição cumulativa, em que os valores de localização estimados foram utilizados como valores de escala para os itens (LINDEN, 2015).

Embora Binet e Thurstone tenham dado importantes contribuições para o início da TRI, foi com o trabalho de Lord e Novick (1968), *Statistical Theories of Mental Tests Scores*, que é dado o início formal da TRI, pois esta obra marca o antes e depois da Teoria dos Testes, ou seja, marca o início da psicometria moderna, a Teoria de Resposta ao Item (MUÑIZ, 1994). Desde então a TRI tem sido muito pesquisada e expandida.

Após uma breve histórico do desenvolvimento da psicometria, buscou-se nas próximas subseções realizar uma breve descrição dos dois modelos, a TCT e TRI, apresentando o modelo matemático e seus pressupostos.

3 Teoria clássica dos testes

O modelo da TCT, ou modelo linear clássico, foi inicialmente desenvolvido por Spearman (MUÑIZ, 1994; PASQUALI, 2009) e axiomatizada por Gulliksen (GULLIKSEN, 1950). Alguns autores também apresentaram um resumo do modelo sistemático do modelo (LORD, 1959; NOVICK, 1966). Com base nesses autores será apresentado os fundamentos da TCT. São três os elementos que constituem o postulado fundamental da teoria:

$$T = V + E$$

Ou seja, o escore empírico é a soma do escore verdadeiro mais o erro, que se define como (GULLIKSEN, 1950):

T = escore bruto ou empírico do sujeito, que é a soma dos escores obtidos no teste;
V = escore verdadeiro, que seria a magnitude real daquilo que o teste quer medir no sujeito e que seria o próprio T se não houvesse erro de medida e;
E = o erro cometido nesta medida.

É importante ressaltar que o erro de medida está presente em qualquer operação empírica, assim, o objetivo da TCT é dispor de técnicas estatísticas que visem controlar ou prever o tamanho do erro na aplicação dos testes (MUÑIZ, 1994; PASQUALI, 2009). Assim, é razoável assumir que erro é definido como a diferença entre o escore verdadeiro (a pontuação real do sujeito) e o escore observado (o escore do sujeito no teste), ou escore empírico (LORD, 1959).

É necessário também destacar que o erro é aleatório, assistemático, isto é, ao realizar várias vezes o teste, o sujeito obterá diferentes pontuações, e essas pontuações poderão ser maiores ou menores que a pontuação verdadeira. Portanto, como o erro é um evento casual, sabe-se que a média do erro é 0 (MUÑIZ, 1994).

São três os pressupostos fundamentais do modelo, dos quais não podem ser empiricamente comprovados de forma direta (MUÑIZ, 1994). O primeiro suposto diz que a pontuação verdadeira (V) é a esperança matemática da pontuação empírica (T): $V = E(T)$. Isso significa que se o teste fosse aplicado infinitas vezes, a pontuação verdadeira seria a média da pontuação empírica. O segundo suposto diz que não há correlação entre a pontuação verdadeira e os erros de um teste, ou seja, a correlação é zero: $\rho(V, E) = 0$. Não há razões para supor que o tamanho da pontuação verdadeira está associado ao tamanho do erro, se assim fosse, o erro seria sistemático, portanto, controlável. O terceiro suposto assume que os erros em dois testes distintos não se correlacionam: $\rho(E_j, E_k) = 0$. Assim, se os testes são aplicados adequadamente, os erros serão aleatórios em cada teste.

Outro aspecto importante na TCT é a definição de testes paralelos. Assumindo que é possível elaborar, diz ser paralelos dois testes que medem a mesma coisa, mas com itens diferentes (MUÑIZ, 1994): são “tau equivalentes” quando possuem pontuações verdadeiras iguais e com variância de erro não necessariamente igual e “essencialmente tau equivalentes” quando as pontuações dos sujeitos são iguais em um e outro teste mais uma constante: $V_1 = V_2 + K$.

No modelo de TCT considera-se principalmente dois parâmetros dos itens, a dificuldade e a discriminação dos itens. Esses dois parâmetros são melhor detalhados na seção a seguir.

3.1 Parâmetros dos itens: dificuldade e discriminação

Para a teoria clássica a dificuldade de um item é definida como a proporção de sujeitos que respondem corretamente ao item (ANDRIOLA, 1998; VIANNA, 1976). A partir disso, um item é mais fácil quando mais pessoas o acertam. Por isso, esse tipo de item, que a resposta pode ser certa ou errada, ou seja dicotômica, é utilizado frequentemente em testes de desempenho ou aptidão em determinada área (PASQUALI, 2009). Por exemplo, um item que apenas 20% dos respondentes o acertam é mais difícil que um item que 80% acertam.

A discriminação, para a teoria clássica, quando entendida no âmbito da avaliação de desempenho escolar, é definida como a capacidade do item distinguir sujeitos de desempenho diferentes, ou seja, aqueles de escores altos em relação àqueles de escores baixos (PASQUALI, 2009; VIANNA, 1976). Este parâmetro também indica a coerência do escore do item com o escore do teste, assim, quanto maior for o índice dos itens, maior será a homogeneidade do teste (SILVEIRA, 1983). O parâmetro de discriminação de um item pode ser obtido pela correlação do item com o escore total menos o escore do item (correlação item-total), que pode ser obtido através da correlação bisserial por ponto. (PASQUALI, 2009).

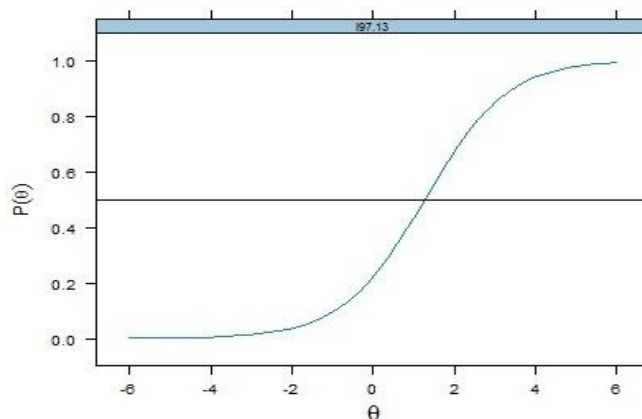
Esses parâmetros dos itens podem ser facilmente observados em contextos de avaliação do ensino e aprendizagem em sala de aula, ou mesmo em avaliações de larga escola e concursos para seleção de candidatos em determinada área. A dificuldade dos itens desejável pode variar de acordo com o objetivo da avaliação. Por exemplo, em contextos de avaliação de sala de aula, é interessante que os itens apresentem dificuldade média, no entanto, em contextos de seleção em que há vários candidatos para poucas vagas, é importante que os itens tenham índice de dificuldade alta, de forma que apenas os candidatos de alto desempenho consigam acertá-lo. Já a discriminação é desejável que o item seja o mais discriminativo quando possível.

4 Teoria de resposta ao item

A TRI surge no âmbito da psicometria moderna como resposta aos problemas apresentados pela TCT e como complemento a esta e não como modelo substitutivo (ANDRIOLA, 2009). Esse modelo trabalha com o traço latente, que é o teta (θ), fenômeno psíquico, como critério a partir dos itens do teste (variáveis observáveis), assim, a qualidade do teste é determinada em função dos itens, com isso, ela objetiva construir itens de qualidade (PASQUALI, 2009), possui foco individualizado nos itens de um teste ou banco de itens (ANDRIOLA, 2009). A TRI sugere que a probabilidade de acertar o item decorra dos seus parâmetros e do traço latente (ou aptidão) do indivíduo exigida em um teste (VALLE, 2000). Dessa forma, esse modelo apresenta uma contribuição importante em relação ao clássico, a possibilidade de comparabilidade dos resultados de indivíduos de uma mesma população, mesmo que submetidos a itens diferentes (MUÑIZ, 1990; PASQUALI, 2009; VALLE, 2000).

Segundo Pasquali (2009), para essa teoria, a probabilidade de acerto a um item aumenta em indivíduos que apresentam maior aptidão e vice-versa. Essa probabilidade ou relação funcional é representada pelo que se denomina de Curva Característica do Item (CCI) (ANDRIOLA, 2009), como mostra a Figura 1.

Figura 1. Curva Característica do Item.



Fonte: Elaboração própria.

Como se pode observar no gráfico, na curva em formato de S ascendente, que é a característica do item, a probabilidade de acertar o item $[P(\theta)]$, na ordenada] aumenta em indivíduos com maior aptidão (θ , na abscissa). Essa curva pode ser afetada por vários parâmetros dependendo do modelo utilizado (MUÑIZ, 1990). Nos

modelos mais usuais de TRI os itens apresentam três parâmetros, dificuldade, discriminação e acerto casual, descritos a seguir.

4.1 Parâmetros da TRI: dificuldade, discriminação e acerto casual

O primeiro parâmetro a ser considerado no item, talvez o mais importante, é o da dificuldade, em que é representado pela letra *b*. Diferentemente da TCT, a TRI o considera na mesma escala do traço latente, ou seja, do teta (θ). Dessa forma, a dificuldade está relacionada ao nível do teta necessário para responder o item (LAROS, 2009).

O índice de discriminação também é um parâmetro que pode ser considerado nos itens. Ele informa a capacidade deste em distinguir sujeitos com habilidades (aptidão) distintas, sendo representado pela letra *a*. Assim, quanto mais o item consegue diferenciar sujeitos com magnitudes próximas de habilidade, mais discriminativo ele é (LAROS, 2009).

Outro parâmetro do item considerado é o acerto ao acaso. Refere-se à probabilidade de acertar o item quando não se tem aptidão suficiente (MUÑIZ, 1990) e é representado pela letra *c*. Como este parâmetro se dá em termos de proporção, ele pode variar de 0 a 1. Quanto maior o valor, maior a probabilidade de os indivíduos acertarem o item ao acaso, dado uma aptidão inferior à dificuldade do item (MUÑIZ, 1990). O modelo mais comum de TRI aplicado na área de avaliação educacional apresenta dois pressupostos, a unidimensionalidade e independência local. Esse pressuposto será descrito com mais detalhes.

4.2 Pressupostos da TRI: unidimensionalidade e independência local

Ao analisar um conjunto de itens a partir da TRI, estes devem dispor de algumas características para que ocorra um adequado ajuste do modelo pretendido. São dois, a unidimensionalidade e independência local. Supõe-se que ao aplicar um teste, ou seja, um conjunto de itens, a probabilidade de acertá-los dependerá unicamente do traço latente do sujeito, do seu θ (MUÑIZ, 1990). Dessa forma, pressupõe-se que os itens estejam medindo um único traço latente, ou seja, que sejam unidimensionais. Este pressuposto “[...] é uma proposição teórica parcimoniosa e elegante, segundo o qual toda a complexidade intrínseca ao ato de resolução de um

problema – de natureza cognitiva ao não – deve-se como causa uma única estrutura latente [...]” (ANDRIOLA, 2009 p. 327). É importante destacar que, há críticas consideráveis a esse pressuposto (TAVARES, 2013), em que aponta-se inconsistência desse pressuposto com a realidade, pois é não é difícil considerar que o conhecimento humano é multideterminado, colocando em dúvidas essa ideia de unidimensionalidade.

Outro pressuposto da TRI é o da independência local. Supõe-se que as respostas a um determinado item não sejam influenciadas por outros itens, ou seja não estão fortemente correlacionados. Se existe independência local, a probabilidade de acertar um conjunto de itens é igual ao produto da probabilidade de acertar cada um destes (MUÑIZ, 1990; PASQUALI, 2009).

No entanto, em contextos práticos da avaliação educacional, tem sido mais comum a verificação do pressuposto da unidimensionalidade. Isso ocorre porque quando um item é unidimensional, ou seja, apresenta alta correlação com o traço latente, sua correlação com outros itens pode ser mais fraca ou não ocorre, atendendo também ao pressuposto da independência local. Assim, garantindo esse pressuposto, três modelos de TRI comumente utilizados em testes educacionais são possíveis de serem aplicados aos dados. Esses modelos são melhor detalhados na seção posterior.

4.3 Modelos de TRI

Existem vários modelos de TRI na literatura e, como coloca Valle (2000), esses dependem fundamentalmente de três fatores: a natureza dos itens (dicotômicos ou não), o número de populações envolvidas e a quantidade de traços latentes medidos (quando mede mais de um denominam-se modelos multidimensionais). Ressalta a autora que os modelos unidimensionais para itens dicotômicos são os mais utilizados. Os modelos logísticos mais comuns são classificados de acordo com os parâmetros utilizados (MUÑIZ, 1990; PASQUALI, 2009; VALLE, 2000): de um parâmetro: dificuldade; de dois parâmetros: dificuldade e discriminação; de três parâmetros: dificuldade, discriminação e probabilidade de acerto ao acaso.

A expressão matemática utilizada para esses modelos são:

Modelo Logístico de 1 parâmetro

$$P_i(\theta) = \frac{e^{D(\theta - b_i)}}{1 + e^{D(\theta - b_i)}}$$

Modelo Logístico de 2 parâmetros

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

Modelo Logístico de 3 parâmetros

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

em que,

$P_i(\theta)$ = é a probabilidade de um sujeito acertar o item dado um determinado teta;

θ = é o nível de habilidade do sujeito;

i = é o número do item no teste;

a = índice de discriminação;

b = índice de dificuldade do item;

c = índice de acerto casual;

e = uma base logarítmica de valor 2,72;

D = é uma constante de valor 1,7.

O modelo de 1 parâmetro utiliza apenas o parâmetro de dificuldade dos itens, ou seja, a resposta ao item depende apenas deste e da aptidão do indivíduo, ou seja, da variável latente (MUÑIZ, 1990). O modelo de 2 parâmetros considera a dificuldade e a discriminação do item, em que a função logística deve considerar estes dois parâmetros (MUÑIZ, 1990; PASQUALI, 2009). O modelo de 3 parâmetros considera a dificuldade, discriminação e a probabilidade de acerto ao acaso (MUÑIZ, 1990; PASQUALI, 2009). Esse modelo de teoria de resposta ao item é o mais utilizado atualmente.

5 Considerações finais

A psicometria ofereceu importantes contribuições para as mais diversas áreas. No campo da avaliação educacional, principalmente com a teoria de resposta ao item, permitiu avanços importantes como a comparação de resultados, mesmo quando utilizados testes distintos e aplicados em populações diferentes. É importante

ressaltar que, embora isso tenha ocorrido, a teoria clássica dos testes não foi simplesmente descartada, ela é ainda muito utilizada e diversos contextos, inclusive o da avaliação, principalmente quando não é possível a aplicação da teoria de resposta ao item. Na verdade, as duas são na maioria das vezes complementares e oferecem informações úteis para análise de instrumentos de medidas educacionais como de seus resultados.

Referências

ANDRIOLA, Wagner Bandeira. Psicometria Moderna: características e tendências. **Estudos em Avaliação Educacional**, v. 20, n. 43, p. 319, 2009. DOI: <https://doi.org/10.18222/eae204320092052>

ANDRIOLA, Waner Bandeira. Avaliação da aprendizagem: uma análise descritiva segundo a teoria de resposta ao item (TRI). **Educação em Debate**, v. 20, n. 36, p. 93–102, 1998. Disponível em: <http://www.periodicosfaced.ufc.br/index.php/educacaoemdebate/article/view/366>
Acesso em: 03 jan. 2020.

ARIAS, Maria Rosario Martínez; LLOREDA, Maria José Hernández; LLOREDA, Maria Victoria Hernández. **Psicometría**. Alianza Ed ed. Madrid, 2014.

GULLIKSEN, Harold. A course in the theory of mental tests. **Psychometrika**, [S. l.], v. 8, n. 4, p. 223–245, 1943. DOI: <https://doi.org/10.1007/BF02288706>

GULLIKSEN, Harold. **Theory of mental tests**. New York: John Wiley & Sons, 1950.

LAROS, J. A. Análise gráfica de itens. In: PASQUALI, L. (org.). **Psicometria: teoria dos testes na psicologia e na educação**. 3. ed. Petrópolis, Rio de Janeiro: Vozes, 2009.

LINDEN, Win J. van der. Introduction. In: LINDEN, Win J. van der (org.). **Handbook of item response theory - volume one**. Boca Raton, FL: CRC Press, 2015.

LORD, Frederic M. An approach to mental test theory. **Psychometrika**, v. 24, n. 4, p. 283–302, 1959. DOI: <https://doi.org/10.1007/BF02289812>

MOREIRA JUNIOR, Fernando de Jesus. Aplicações da Teoria da Resposta ao Item (TRI) no Brasil. **Revista Brasileira De Biometria**, v. 28, n. 4, p. 137–170, 2010.

MUÑIZ, José. **Teoría de repuesta a los ítems**. Madrid: Pirámede, 1990.

MUÑIZ, José. **Teoría clásica de los testes**. Madrid: Pirámede, 1994.

NOVICK, Melvin R. The axioms and principal results of classical test theory. **Journal**

of **Mathematical Psychology**, v. 3, n. 1, p. 1–18, 1966. DOI: [https://doi.org/10.1016/0022-2496\(66\)90002-2](https://doi.org/10.1016/0022-2496(66)90002-2)

PASQUALI, Luiz. **Psicometria: teoria dos testes na psicologia e na educação**. 3. ed. Petrópolis, Rio de Janeiro: Vozes, 2009.

SARTES, Laisa Marcorela Andreoli; SOUSA-FORMIGONI, Maria Lucia Oliveira de. Avanços na Psicometria: da teoria clássica dos testes à teoria de resposta ao item. **Psicologia: Reflexão e Crítica**, v. 26, n. 2, p. 241–250, 2013. DOI: <https://doi.org/10.1590/S0102-79722013000200004>

SILVEIRA, Fernando Lang da. Considerações sobre o índice de discriminação de itens em testes educacionais. **Educação & Seleção**, v. 7, 1983. Disponível em: <http://publicacoes.fcc.org.br/ojs/index.php/edusel/article/view/2546> Acesso em: 06 jan. 2020

SOUSA, Leandro Araujo de et al. Teoria de resposta ao item como possibilidade de análise de testes educacionais de educação física. In: PONTES JUNIOR, José Aírton de Freitas et al. (org.). **Métodos quantitativos em atividade física**. Curitiba: CRV, 2018.

SOUSA, Leandro Araujo de; BRAGA, Adriana Eufrásio; TROMPIERI FILHO, Nicolino. Educação Física em avaliações de larga escala: o caso do exame nacional do ensino médio - Enem. **Revista educação & Linguagem**, v. 2, n. 1, p. 60–71, 2015. Disponível em: <https://www.fvj.br/revista/revista-educacao-e-linguagem/edicoes/2015-1/> Acesso em: 06 jan. 2020

STEVENS, S. S. On the theory of scales of measurement. **Science**, v. 103, n. 2684, p. 677–680, 1946. DOI: <https://doi.org/10.1126/science.103.2684.677>

TAVARES, Cristina Zukowsky. Teoria da resposta ao item: uma análise crítica dos pressupostos epistemológicos. **Estudos em Avaliação Educacional**, v. 24, n. 54, p. 56–76, 2013. DOI: <http://dx.doi.org/10.18222/eae245420131902>

TRAUB, Ross E. Classical test theory in historical perspective. **Educational Measurement: Issues and Practice**, p. 8–14, 1997. DOI: <https://doi.org/10.1111/j.1745-3992.1997.tb00603.x>

URBINA, Susana. **Essentials of psychological testing**. New Jersey: John Wiley & Sons, Inc., 2004.

VALLE, Raquel da Cunha. Teoria de resposta ao item. **Estudos em Avaliação educacional**, v. 21, p. 07–92, 2000. DOI: <http://dx.doi.org/10.18222/eae02120002225>

VIANNA, Heraldo Marelím. **Testes em educação**. São Paulo: IBRASA, 1976.

i **Leandro Araujo de Sousa**, ORCID: <https://orcid.org/0000-0002-0482-2699>

Doutor em educação pela Universidade Federal do Ceará (UFC). Docente do Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE).

Contribuição de autoria: Conceituação, Primeira Redação.

Lattes: <http://lattes.cnpq.br/7393610074936097>.

E-mail: leandro.sousa@ifce.edu.br

ii **Adriana Eufrásio Braga**, ORCID: <https://orcid.org/0000-0001-5163-209X>

Doutora em educação pela Universidade Federal do Ceará (UFC). Docente do Programa de Pós-Graduação em Educação (PPGE/UFC).

Contribuição de autoria: Supervisão, Escrita – Revisão e Edição.

Lattes: <http://lattes.cnpq.br/8042943225979510>.

E-mail: adrianaufc@yahoo.com.br